

DOCUMENT RESUME

ED 167 566

TH 007 657

AUTHOR Folsom, Ralph E., Jr.
TITLE A Survey Measurement Error Model for Repeated Binary Responses.
PUB DATE [75]
NOTE 24p.
EDRS PRICE MF-\$0.83 HC-\$1.67 Plus Postage.
DESCRIPTORS *Longitudinal Studies; Reliability; *Research Design; Standard Error of Measurement; *Statistical Analysis; *Statistical Bias; *Surveys
IDENTIFIERS *Repeated Measures Designs

ABSTRACT

In large-scale surveys, it is no longer uncommon for repeated measurements to be obtained from respondents and analyses performed to gauge the magnitude of nonsampling errors. This is particularly true for periodic surveys and longitudinal surveys where a very large investment in data collection is made. This technical note, aimed at the analysis of repeated survey measurements, emphasizes the misclassification error aspects of the Census Bureau model for zero-one variables. Section 2 of this note presents a specification of the Census Bureau model for zero-one variables which postulates a "true value" for each potential respondent and emphasizes the misclassification aspects of the error distribution. Section 3 explores the bias and variance implications of the measurement error model for sample proportions. Section 4 presents the expectations of several measures of inconsistency introduced for repeated measurement surveys by Hansen, Hurquitz, and Pritzker. Section 5 summarizes the results, concentrating on their practical implications regarding the analysis of repeated survey measurements. (Author/CTH)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Research Triangle Inst.

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC) AND
USERS OF THE ERIC SYSTEM."

A SURVEY MEASUREMENT ERROR MODEL FOR REPEATED BINARY RESPONSES

by

Ralph E. Folsom, Jr.

1. Introduction

Since the publication of Hansen, Hurwitz, and Berstad's [1] 1959 paper on "Measurement Errors in Census and Surveys," sampling statisticians have paid increasing attention in their methodological research and survey applications to nonsampling errors. In large-scale surveys, it is no longer uncommon for repeated measurements to be obtained from respondents and analyses performed to gauge the magnitude of nonsampling errors. This is particularly true for periodic surveys and longitudinal surveys where a very large investment in data collection will be made. This technical note aimed at the analysis of repeated survey measurements emphasizes the misclassification error aspects of the Census Bureau model for zero-one variables. While the Hansen, et.al paper concentrates on the zero-one case, the general flavor of their results is the familiar random effects one-way analysis of variance (ANOVA) model allowing for correlations among the measurement errors. While this ANOVA approach neatly partitions the total variance of sample statistics in terms of variance and covariance components for both sampling and measurement sources, it tends to downplay the biasing effects of nonsampling errors. For binary (zero-one) responses where there can be no measurement source of variance without associated

ED167566

IM007 657

measurement bias, a more detailed specification of the Census Bureau model seems indicated.

Section 2 of this note presents a specification of the Census Bureau model for zero-one variables which postulates a "true value" for each potential respondent and emphasizes the misclassification aspects of the error distribution. Section 3 explores the bias and variance implications of the measurement error model for sample proportions. Section 4 presents the expectations of several measures of inconsistency introduced for repeated measurement surveys by Hansen, Hurwitz, and Pritzker [2]. Section 5 summarizes the results concentrating on the practical implications regarding the analysis of repeated survey measurements.

2. Measurement Error Model

This note will focus on a finite population with N members. A simple random sample of n population members is interviewed. For a particular question with two possible responses coded one or zero, the sample members are presumed to have true response values

$X(i) \equiv$ the true zero or one response for sample member- i .

The corresponding true proportion of population members with $X = 1$ responses is

$$\bar{X} = \frac{1}{N} \sum_{k=1}^N X(k)/N = \pi. \quad (2.1)$$

The N population members in the sum above have been indexed by k to distinguish them from the sample members. The population variance for

the true X's is

$$\begin{aligned}\sigma_X^2 &= \sum_{k=1}^N [X(k) - \bar{X}]^2 / N \\ &= \pi(1 - \pi) .\end{aligned}\tag{2.2}$$

For a simple random sample of size n selected without replacement,
the sample mean

$$\bar{x} = \sum_{i=1}^n X(i) / n$$

has expectation over repeated samples of size n

$$E_s(\bar{x}) = \pi$$

and variance

$$\text{Var}_s(\bar{x}) = [1 - (n-1)/(N-1)] \pi(1 - \pi) / n ,$$

To account for measurement or response errors, Hansen, et.al postulate a survey procedure--unattainable in the real world--in which the process of recording a response for any sample member is repeatable under the same general survey conditions, and that these repeated measurements will be essentially uncorrelated. With this assumption, one can model the response from sample member- i on resurvey trial- t under general survey conditions- G as

$$x_G(it) = X(i) + d_G(it)\tag{2.3}$$

where $d_G(it) = [x_G(it) - X(i)]$ denotes the measurement or response error.

For 1-0 variables the response deviations $d_G(it)$ above have a particular form, namely

$$d_G(it) = X(i) d_G(it|1) + [1 - X(i)] d_G(it|0) \quad (2.4)$$

with

$$d_G(it|1) = \begin{cases} -1 & \text{with probability } P_G(i|1) \text{ over trials-}t \\ 0 & \text{with probability } [1 - P_G(i|1)] \end{cases}$$

and

$$d_G(it|0) = \begin{cases} 1 & \text{with probability } P_G(i|0) \text{ over trials-}t \\ 0 & \text{with probability } [1 - P_G(i|0)] . \end{cases}$$

Given the true value of $X(i)$, it is apparent that the response deviation may take only those values prescribed in equation 2.4. With the assumption of uncorrelated errors over trials, one notes that the conditional response deviations given the true $X(i)$ values are distributed as independent Bernoulli trials. The expected value of $d_G(it)$ over trials is therefore

$$E_t\{d_G(it)\} = -X(i) P_G(i|1) + [1 - X(i)] P_G(i|0) . \quad (2.5)$$

Readers familiar with misclassification error models will recognize the probabilities $P_G(i|1)$ and $P_G(i|0)$ in equations 2.4 and 2.5 as the probabilities of "false negative" and "false positive" responses, respectively, where $x_G(it) = 1$ corresponds to a Yes response at trial- t . The

variance of $d_G(it)$ over trials is

$$\begin{aligned} \text{Var}_t\{d_G(it)\} = \sigma_{dG}^2(i) = & X(i) P_G(i|1)[1 - P_G(i|1)] \\ & + [1 - X(i)] P_G(i|0)[1 - P_G(i|0)]. \quad (2.6) \end{aligned}$$

3. Bias and Variance Implications

If the measurement process is to be unbiased for sample member- i , then the expectation in equation 2.5 must equal zero. If the true response for sample member- i is $X(i) = 1$, then $E_t\{d_G(it)\} = 0$ requires that $P_G(i|1) = 0$. Alternatively, if $X(i) = 0$, then unbiasedness requires that $P_G(i|0) = 0$. In either of these instances unbiasedness translates into no chance of misclassification over trials; that is, $E_t\{d_G(it)\} = 0$ implies $d_G(it) = 0$ for all trials- t . For a binary response, the immediate consequence of an unbiased measurement is an error free measurement. Therefore, one may not legitimately assume that $E_t\{d_G(it)\} = 0$ without accepting the corollary that $\sigma_{dG}^2(i) = 0$.

Allowing all the individual response biases characterized by equation 2.5 to be nonzero, the net bias in the overall sample mean $\bar{x}_G(t)$ is

$$\text{Bias}\{\bar{x}_G(t)\} = E_s E_t \{\bar{d}_G(t)\}$$

where E_s denotes expectation over repeated samples of size n . Applying the result from equation 2.5, one observes that

$$\text{Bias}\{\bar{x}_G(t)\} = E_s \left\{ - \sum_{i=1}^n X(i) P_G(i|1)/n + \sum_{i=1}^n [1 - X(i)] P_G(i|0)/n \right\}$$

$$\begin{aligned}
&= - \sum_{k=1}^N X(k) P_G(k|1)/N + \sum_{k=1}^N [1 - X(k)] P_G(k|0)/N \\
&= - \pi \bar{P}_G(0|1) + (1 - \pi) \bar{P}_G(1|0)
\end{aligned} \tag{3.1}$$

where the P parameters in equation 3.1 represent population averages of the individual misclassification probabilities. If $N(1) = \sum_{k=1}^N X(k)$ denotes the "true" number of population members with $X(k) = 1$, then

$$\bar{P}_G(0|1) = \sum_{k \in 1} P_G(k|1)/N(1)$$

represents the average probability of a false negative response. Similarly,

$$\bar{P}_G(1|0) = \sum_{k \in 0} P_G(k|0)/[N - N(1)]$$

depicts the average probability of a false positive response. If the net bias in $\bar{x}_G(t)$ is to be zero, then equation 3.1 requires that

$$\bar{P}_G(1|0) = \pi \bar{P}_G(0|1)/(1 - \pi) . \tag{3.2}$$

While it is therefore technically possible to have biased individual measurements which when averaged together yield a zero net bias, finding a population and a set of survey conditions where some of the individual misclassification probabilities are nonzero and the relationship in equation 3.2 holds seems extremely unlikely.

Reviewing the results stated in the previous paragraphs, it has been shown for zero-one variables that where measurement errors are anticipated, nonsampling bias as well as measurement variance must be expected. Denoting the individual measurement bias in equation 2.5

by $B_G(i)$, the net or average bias in equation 3.1 can be represented by

$$\bar{B}(G) = \sum_{k=1}^N B_G(k)/N = -\pi \bar{P}_G(0|1) + (1-\pi) \bar{P}_G(1|0) \quad (3.3)$$

To evaluate the variance of $\bar{x}_G(t)$, the following partitioning

$$\text{Var}\{\bar{x}_G(t)\} = \text{Var}_s\{E_t[\bar{x}_G(t)]\} + E_s\{\text{Var}_t[\bar{x}_G(t)]\} \quad (3.4)$$

is useful. The first term in equation 3.4 is

$$\text{Var}_s\left\{\sum_{i=1}^n [X(i) + B_G(i)]/n\right\} = [1 - (n-1)/(N-1)] [\sigma_X^2 + \sigma_B^2(G) + 2\sigma_{XB}(G)]/n$$

where

$$\sigma_X^2 = \pi(1-\pi) \quad (3.5)$$

is the sampling variance among the true X's,

$$\sigma_B^2(G) = \pi \sigma_G^2(0|1) + (1-\pi) \sigma_G^2(1|0) + \pi(1-\pi) [\bar{P}_G(0|1) + \bar{P}_G(1|0)]^2$$

with

$$\sigma_G^2(0|1) = \sum_{k \in 1} [P_G(k|1) - \bar{P}_G(0|1)]^2 / N(1)$$

and

$$\sigma_G^2(1|0) = \sum_{k \in 0} [P_G(k|0) - \bar{P}_G(1|0)]^2 / [N - N(1)]$$

is the sampling variance among the individual measurement biases, and

$$\sigma_{XB}(G) = -\pi(1-\pi) [\bar{P}_G(0|1) + \bar{P}_G(1|0)]$$

is the sampling covariance among the true X's and the individual measurement biases. Writing equation 3.5 in terms of the misclassification model parameters yields

$$\begin{aligned} \text{Var}_s E_t [\bar{x}_G(t)] &= [1 - (n-1)/(N-1)] \pi(1-\pi) \{ [1 - \bar{P}_G(+)]^2 \\ &\quad + \sigma_G^2 (0|1)/(1-\pi) + \sigma_G^2 (1|0)/\pi \} / n \end{aligned}$$

where

$$\bar{P}_G(+) = \bar{P}_G(0|1) + \bar{P}_G(1|0) .$$

The second term in equation 3.4 is evaluated subject to the restriction that

$$E_t \{ [d_G(it) - B_G(i)] [d_G(i't) - B_G(i')] \} = 0$$

for all (i, i') belonging to the sample or equivalently that the measurement errors for different sample members are uncorrelated over trials. This restriction which is best justified for self-administered mail surveys yields

$$\begin{aligned} E_s \left\{ \sum_{i=1}^n \text{Var}_t [d_G(it)] / n^2 \right\} &= \sum_{k=1}^N \sigma_{dG}^2(k) / n N \\ &= \sigma_d^2 / n \end{aligned} \tag{3.6}$$

with

$$\begin{aligned} \sigma_d^2 &= \pi \sum_{k \in 1} P_G(k|1) [1 - P_G(k|1)] / N(1) \\ &\quad + (1 - \pi) \sum_{k \in 0} P_G(k|0) [1 - P_G(k|0)] / [N - N(1)] \end{aligned}$$

$$\begin{aligned}
&= \pi \bar{P}_G(0|1) [1 - \bar{P}_G(0|1)] + (1 - \pi) \bar{P}_G(1|0) [1 - \bar{P}_G(1|0)] \\
&\quad - \pi \sigma_G^2(0|1) - (1 - \pi) \sigma_G^2(1|0)
\end{aligned}$$

representing the simple response or measurement variance component. For with replacement selections or cases where the finite population correction can be ignored, the total variance of $\bar{x}_G(t)$ is

$$\begin{aligned}
\text{Var}(\bar{x}_G(t)) &= [\pi(1-\rho) + \sigma_B^2 + 2\sigma_{XB} + \sigma_d^2]/n \\
&= [\pi + \bar{B}(G)] [1 - \pi - \bar{B}(G)]/n \\
&= \pi(G) [1 - \pi(G)]/n \tag{3.7}
\end{aligned}$$

where

$$\begin{aligned}
\pi(G) &= E_s E_t \{\bar{x}_B(t)\} = \pi + \bar{B}(G) \\
&= \pi[1 - \bar{P}_G(0|1)] + (1 - \pi) [1 - \bar{P}_G(1|0)] .
\end{aligned}$$

Equation 3.7 demonstrates the fact that for with replacement selections, $\bar{x}_G(t)$ has the sampling distribution of a binomial proportion with expectation $\pi(G) = \pi + \bar{B}(G)$. ✓

Having characterized the bias and variance implications of the measurement error model introduced in section 2 for a sample proportion $\bar{x}_G(t) = \hat{\pi}(G)$, we now turn to the analysis of a second set of measurements on the initial sample of n members. It is customary to assume that the second set of measurements are obtained under identical survey conditions-G and are independent of, or uncorrelated with, the first.

These second measurements will be denoted by

$$x_{G'}(it') = X(i) + d_{G'}(it') . \quad (3.8)$$

The distribution over repeated trials of the measurement errors above will be characterized by misclassification probabilities $P_{G'}(i|1)$ or $P_{G'}(i|0)$ depending on whether $X(i) = 1$ or 0 . Paralleling the results obtained for the first measurement, one can show that

$$E_s E_{t'} \{\bar{x}_{G'}(t')\} = \pi + \bar{B}(G') = \pi(G') \quad (3.9)$$

with

$$\bar{B}(G') = -\pi \bar{P}_{G'}(0|1) + (1 - \pi) \bar{P}_{G'}(1|0) .$$

For an initial with replacement selection,

$$\text{Var}\{\bar{x}_{G'}(t')\} = \pi(G')[1 - \pi(G')]/n \quad (3.10)$$

represents the total variance of the estimated proportion based on second measurements. To account for correlations between the first and second measurement errors, let

$$P_{GG'}(i|1) = \Pr_{t,t'} \{d_G(it|1) = d_{G'}(it'|1) = -1\}$$

and

$$P_{GG'}(i|0) = \Pr_{t,t'} \{d_G(it|0) = d_{G'}(it'|1) = 1\}$$

denote the probabilities that sample member- i is misclassified on both occasions depending on whether $X(i) = 1$ or 0 . In terms of these double misclassification probabilities, the covariance over trials- t and t'

of the first and second measurement errors is

$$\begin{aligned} \text{Cov}_{t,t'}\{d_G(it); d_G(it')\} &= X(i)[P_{GG'}(i|1) - P_G(i|1)P_{G'}(i|1)] \\ &\quad + [1 - X(i)][P_{GG'}(i|0) - P_G(i|0)P_{G'}(i|0)] \\ &= \sigma_{dd'}(i) . \end{aligned} \quad (3.11)$$

Evaluating the covariance between the sample proportions $\bar{x}_G(t)$ and $\bar{x}_{G'}(t')$ based on the first and second measurements, the following partitioning is exploited.

$$\begin{aligned} \text{Cov}(\bar{x}_G(t); \bar{x}_{G'}(t')) &= \text{Cov}_S\{E_t[\bar{x}_G(t)]; E_{t'}[\bar{x}_{G'}(t')]\} \\ &\quad + E_S \text{Cov}_{t,t'}\{\bar{x}_G(t); \bar{x}_{G'}(t')\} . \end{aligned} \quad (3.12)$$

The first term in equation 3.12 above becomes

$$\begin{aligned} \text{Cov}_S\left\{\sum_{i=1}^n [X(i) + B_G(i)]/n; \sum_{i=1}^n [X(i) + B_{G'}(i)]/n\right\} \\ = [1 - (n-1)/(N-1)] (\sigma_X^2 + \sigma_{XB} + \sigma_{XB'} + \sigma_{BB'})/n \end{aligned} \quad (3.13)$$

where σ_{XB} has the same form as σ_{XB} in equation 3.5 with G' error probabilities replacing the corresponding G parameters. The sample covariance among the individual measurement biases can be expanded to

$$\begin{aligned} \sigma_{BB'} &= \pi \sum_{k \in 1} P_G(k|1)P_{G'}(k|1)/N(1) \\ &\quad + (1-\pi) \sum_{k \in 0} P_G(k|0)P_{G'}(k|0)/[N - N(1)] \end{aligned}$$

$$= -\pi^2 \bar{P}_G(0|1) \bar{P}_G(0|1) - (1-\pi^2) \bar{P}_G(1|0) \bar{P}_G(1|0) \\ + \pi(1-\pi) [\bar{P}_G(0|1) \bar{P}_G(1|0) + \bar{P}_G(1|0) \bar{P}_G(0|1)] .$$

With some further manipulation, σ_{BB} , can be expressed as follows

$$\sigma_{BB} = \pi \sigma_{GG}(0|1) + (1-\pi) \sigma_{GG}(1|0) \\ + (1-\pi) [\bar{P}_G(0|1) + \bar{P}_G(1|0)] [\bar{P}_G(0|1) + \bar{P}_G(1|0)] \quad (3.14)$$

where

$$\sigma_{GG}(0|1) = \sum_{k \in 1} [P_G(k|1) - \bar{P}_G(0|1)] [P_G(k|1) - \bar{P}_G(0|1)] / N(1)$$

and $\sigma_{GG}(1|0)$ is the corresponding covariance between the first and second measurement error probabilities for population members whose true responses are $X(k) = 0$. In terms of the misclassification model parameters, one notes that the sampling covariance portion of equation 3.12 becomes

$$\text{Cov}_s \{E_t[\bar{x}_G(t)]; E_t[\bar{x}_G(t)]\} \\ = [1 - (n-1)/(N-1)] \pi(1-\pi) \{[1 - \bar{P}_G(+)] [1 - \bar{P}_G(+)] \\ + \sigma_{GG}(0|1)/(1-\pi) + \sigma_{GG}(1|0)/\pi\} / n$$

where

$$\bar{P}_G(+) = \bar{P}_G(0|1) + \bar{P}_G(1|0) .$$

Turning to the second term in equation 3.12, the measurement covariance term, the restriction that between individual covariances of the form

$$\text{Cov}_{t,t'}\{d_G(it); d_G(i't')\} = 0$$

leads to the result

$$\begin{aligned} E_s \text{Cov}_{t,t'}\{\bar{x}_G(t); \bar{x}_G(i')\} &= E_s \left\{ \sum_{k=1}^n \sigma_{dd'}(1)/n^2 \right\} \\ &= \{\pi[\bar{P}_{GG'}(00|1) - \bar{P}_G(0|1)\bar{P}_{G'}(1|0)] \\ &\quad + (1-\pi)[\bar{P}_{GG'}(11|0) - \bar{P}_G(1|0)\bar{P}_{G'}(1|0)] \\ &\quad - \pi \sigma_{GG'}(0|1) - (1-\pi)\sigma_{GG'}(1|0)\}/n \\ &= \sigma_{dd'}/n \end{aligned} \quad (3.15)$$

For an initial with replacement sample of n population members, the overall covariance expression in equation 3.12 becomes

$$\begin{aligned} \text{Cov}\{\bar{x}_G(t); \bar{x}_{G'}(t)\} &= (\sigma_X^2 + \sigma_{XB} + \sigma_{XB'} + \sigma_{BB'} + \sigma_{dd'})/n \\ &= [\pi(GG') - \pi(G)\pi(G')]/n \end{aligned} \quad (3.16)$$

where the $\pi(G)$ and $\pi(G')$ parameters were defined previously, and

$$\pi(GG') = \pi[1 - \bar{P}_G(0|1) - \bar{P}_{G'}(0|1) + \bar{P}_{GG'}(00|1)] + (1-\pi)\bar{P}_{GG'}(11|0)$$

represents the overall probability that a population member selected

by simple random sampling with replacement would respond Yes or

$X = 1$ on both occasions.

4.0 Expectations of NDR, GRR, and $R = (1 - I)$

In the second Hansen, et.al publication [2] referenced previously, three statistics were recommended for summarizing repeated binary responses. The first of these statistics was called the net difference rate (NDR). The net difference rate is simply the difference between the two sample means $\bar{x}_G(t)$ and $\bar{x}_{G'}(t)$ based on the first and second measurements respectively. Emphasizing the fact that these means represent estimates of the $\pi(G)$ and the $\pi(G')$ proportions, NDR is simply

$$NDR = \hat{\pi}(G) - \hat{\pi}(G') \quad (4.1)$$

With reference to the misclassification error model developed in section 2, NDR is seen to estimate

$$\begin{aligned} E(NDR) &= \pi(G) - \pi(G') \\ &= \bar{B}(G) - \bar{B}(G') \\ &= -\pi[\bar{P}_G(0|1) - \bar{P}_{G'}(0|1)] + (1-\pi)[\bar{P}_G(1|0) - \bar{P}_{G'}(1|0)] \end{aligned} \quad (4.2)$$

Equation 4.2 demonstrates that NDR is estimating the difference between the net biases at measurement occasions one and two. One notes that for $E(NDR)$ to be zero, either the average misclassification probabilities for $X = 1$ and $X = 0$ population members must be equal on both occasions or

$$[\bar{P}_G(0|1) - \bar{P}_{G'}(0|1)] = (1 - \pi) [\bar{P}_G(1|0) - \bar{P}_{G'}(1|0)]/\pi$$

For the second condition yielding $E(NDR) = 0$ to be satisfied appears such an unlikely event one could probably limit the interpretation of zero NDR to the equivalence of average misclassification probabilities. An approximate chi-square statistic with one degree of freedom for testing the hypothesis that $NDR = 0$ is

$$q = (NDR)^2 / \text{var}(NDR) \quad (4.3)$$

The estimated variance for NDR in the denominator of q above has the form

$$\begin{aligned} \text{var}(NDR) = & \{ \hat{\pi}(G)[1 - \hat{\pi}(G)] + \hat{\pi}(G')[1 - \hat{\pi}(G')] \\ & - 2[\hat{\pi}(GG') - \hat{\pi}(G)\hat{\pi}(G')] \} / n \end{aligned}$$

where $\hat{\pi}(GG')$ represents the fraction of sample members who respond Yes or 1 on both occasions. With binary (zero-one) responses, q is equivalent to the statistic proposed by Bhapkar [3] for testing marginal homogeneity in the 2×2 table displayed in figure 1:

		<u>Second Measurement</u>		<u>Total</u>
		<u>Yes</u>	<u>No</u>	
<u>First Measurement</u>	Yes	$\hat{\pi}(GG')$	$\hat{\pi}(G) - \hat{\pi}(GG')$	$\hat{\pi}(G)$
	No	$\hat{\pi}(G') - \hat{\pi}(GG')$	$1 - \hat{\pi}(G) - \hat{\pi}(G') + \hat{\pi}(GG')$	$1 - \hat{\pi}(G)$
Total		$\hat{\pi}(G')$	$1 - \hat{\pi}(G')$	1

Figure 1

The "gross difference rate" defined by Hansen, Hurwitz, and Pritzker

as

$$GDR = \frac{1}{n} \sum_{i=1}^n [x_G(it) - x_{G'}(it')]^2 / n \quad (4.4)$$

estimates

$$\begin{aligned}
 E(\text{GDR}) &= \sigma_d^2 + \sigma_{d'}^2 - 2\sigma_{dd'} + \frac{1}{N} \sum_{k=1}^N [B_G(k) - B_{G'}(k)]^2 \\
 &= \sigma_d^2 + \sigma_{d'}^2 - 2\sigma_{dd'} + \sigma_B^2 + \sigma_{B'}^2 - 2\sigma_{BB'} \\
 &\quad + [\bar{B}(G) - \bar{B}(G')]^2
 \end{aligned} \tag{4.5}$$

Recalling the results of equations 3.7, 3.16 and 4.2, one notes that for with replacement sampling

$$E(\text{GDR}) = n \text{Var}(\text{NDR}) + E^2(\text{NDR})$$

which translates into

$$E(\text{GDR}) = \pi(G) + \pi(G') - 2\pi(GG') \tag{4.6}$$

In terms of the 2 x 2 table of sample proportions displayed in figure 1, one observes that

$$\text{GDR} = \hat{\pi}(G) + \hat{\pi}(G') - 2\hat{\pi}(GG')$$

which represents the proportion of sample members whose responses changed between the first and second measurements or correspondingly, the sum of off diagonal proportions in figure 1. Writing the π parameters in equation 4.6 in terms of the misclassification error parameters, one obtains

$$\begin{aligned}
 E(\text{GDR}) &= \pi[\bar{P}_G(0|1) + \bar{P}_{G'}(0|1) - 2\bar{P}_{GG'}(00|1)] \\
 &\quad + (1 - \pi) [\bar{P}_G(1|0) + \bar{P}_{G'}(1|0) - 2\bar{P}_{GG'}(11|0)] \tag{4.7}
 \end{aligned}$$

Notice that

$$[P_G(i|1) + P_{G'}(i|1) - 2P_{GG'}(i|1)] = \Pr\{[x_G(it) - x_{G'}(it')]^2 = 1 | X(i) = 1\}$$

represents the non-negative probability that sample member- i will respond (Yes = 1, No = 0) or (No = 0, Yes = 1) on the two occasions, given that his true response is Yes or $X(i) = 1$. Averaging these bivariate response probabilities over the $N(1)$ population members with $X(i) = 1$ illustrates that

$$[\bar{P}_G(0|1) + \bar{P}_{G'}(0|1) - 2\bar{P}_{GG'}(00|1)] = [\bar{P}_{GG'}(10|1) + \bar{P}_{GG'}(01|1)]$$

and by analogy

$$[\bar{P}_G(1|0) + \bar{P}_{G'}(1|0) - 2\bar{P}_{GG'}(11|1)] = [\bar{P}_{GG'}(10|0) + \bar{P}_{GG'}(01|0)]$$

(4.8)

where

$$\bar{P}_{GG'}(rs|u) = \sum_{k \in u} \Pr\{x_G(kt) = r \text{ and } x_{G'}(kt') = s | X(k) = u\} / N(u)$$

with the summation $k \in u$ denoting the sum over all population members- k with $X(k) = u$ and $N(u)$ representing the total number of members with $X(k) = u$. For GDR to be zero requires that the two response probability averages in equation 4.8 are zero, which in turn requires that

$$\Pr\{[x_G(kt), x_{G'}(kt')] = (1,1) \text{ or } (0,0)\} = 1 \quad (4.9)$$

for all population members- k . As a direct consequence of equation 4.9, one can conclude that measurement errors $d_G(kt)$ and $d_{G'}(kt')$ obtained on the two occasions are perfectly correlated for all population members.

For making statistical inference regarding the significance of GDR or constructing confidence intervals about GDR, the with replacement variance approximation is

$$\text{var}(\text{GDR}) = \text{GDR} (1 - \text{GDR})/n \quad (4.10)$$

recognizing that GDR is distributed as a binomial proportion when the initial sample is simple random with replacement. The complementary proportion $(1 - \text{GDR})$ estimates for the with replacement case the probability of agreement between first and second measurements for a randomly selected population member. That is, if $(1 - \text{GDR}) = \text{POA}$ estimates the probability of agreement, then

$$\begin{aligned} E\{\text{POA}\} &= \pi [\bar{P}_{GG}, (11|1) + \bar{P}_{GG}, (00|1)] \\ &+ (1 - \pi) [\bar{P}_{GG}, (11|0) + \bar{P}_{GG}, (00|1)] \end{aligned} \quad (4.11)$$

and values of POA near 1 indicate high positive correlation between the response errors made on the two measurement occasions. The with replacement variance approximation for POA is equivalent to that for GDR quoted in equation 4.10.

The third statistic recommended in reference 2 is the Index of Inconsistency. This index is designed to measure the fraction of total variance for a sample mean $\bar{x}_G(t)$ that can be attributed to the simple uncorrelated response or measurement variance component σ_d^2 . For a with replacement equal probability selection where the measurement errors $d_G(it)$ are uncorrelated for different sample members i and i' , we have shown that a sample mean $\bar{x}_G(t)$ or specifically a sample proportion $\hat{\pi}(G)$ has

total variance

$$\begin{aligned}\text{Var}\{\hat{\pi}(G)\} &= [\pi(1 - \pi) + \sigma_B^2 + 2\sigma_{XB}]/n \\ &\quad + \sigma_d^2/n \\ &= \pi(G) [1 - \pi(G)]/n\end{aligned}$$

with the first quantity in brackets above representing the sampling variance component. If one further assumes that the response errors $d_G(it)$ and $d_G(it')$ obtained on two occasions are independent and identically distributed at least to the extent that the corresponding response biases $B_G(i)$ and $B_G(i')$ along with the associated response variances $\sigma_{dG}^2(i)$ and $\sigma_{dG}^2(i')$ are equal, then from equation 4.5 one notes that

$$E(\text{GDR}) = 2\sigma_d^2 \quad (4.12)$$

Furthermore, with the response biases equal on the two occasions, a combined estimate of the common total variance component is

$$\text{TVC} = \{\hat{\pi}(G)[1 - \hat{\pi}(G)] + \hat{\pi}(G')[1 - \hat{\pi}(G')]\}/2 \quad (4.13)$$

Recalling from equation 4.6 that $\text{GDR} = \hat{\pi}(G) + \hat{\pi}(G') - 2\hat{\pi}(GG')$, an estimate for the Index of Inconsistency $I = \sigma_d^2/\pi(G)[1 - \pi(G)]$ suggested in reference 2 is

$$\begin{aligned}\hat{I} &= \text{GDR}/2\text{TVC} \\ &= [\hat{\pi}(G) + \hat{\pi}(G') - 2\hat{\pi}(GG')]/\{\hat{\pi}(G)[1 - \hat{\pi}(G)] \\ &\quad + \hat{\pi}(G')[1 - \hat{\pi}(G')]\}.\end{aligned}$$

With the estimator for I proposed above, it is clear that small values of GDR will lead to small values of $\hat{I}$ and that the interpretation of small GDR in terms of high correlation between first and second measurements can be extended likewise to small $\hat{I}$ values. Another way of obtaining this correlation interpretation is to recast $\hat{I}$ in terms of a correlation type statistic

$$\hat{\delta} = [\hat{\pi}(GG') - \hat{\pi}(G) \hat{\pi}(G')]/TVC. \quad (4.14)$$

Notice that with $\pi(G)$ and $\pi(G')$ equal,

$$E(TVC) = \pi(G)[1 - \pi(G)] = \{\pi(G)[1 - \pi(G)]\pi(G')[1 - \pi(G')]\}^{1/2}$$

and $\hat{\delta}$ represents a consistent estimate of the product moment correlation between the first and second measurements of a randomly selected population member. Recasting $\hat{I}$ in terms of $\hat{\delta}$ and the single degree of freedom chi-squared statistic q for testing $NDR = 0$, one observes that

$$\begin{aligned} \hat{I} &= [n \text{ var}\{NDR\} + NDR^2]/2 \text{ TVC} \\ &= n \text{ var}\{NDR\} (1 + q/n)/2 \text{ TVC} \\ &= (1 - \hat{\delta}) (1 + q/n). \end{aligned} \quad (4.15)$$

Therefore, if NDR is insignificant or more specifically, if $q(1 - \hat{\delta})/n$ is negligible, then the index of reliability defined by

$$\hat{R} = (1 - \hat{I}) = \hat{\delta} - q(1 - \hat{\delta})/n$$

may be approximated closely by $\hat{\delta}$ which will in turn closely approximate the product moment correlation between first and second measurements.

5. Implications for Analysis of Repeated Survey Measurements

Hansen, Hurwitz, and Pritzker in reference 2 provide the following admonition:

"While measurement of accuracy should be an important guide to the users of data, the education of users in the proper interpretation of accuracy measurements has not been sufficiently effective. Continued effort in this type of educational work is sorely needed."

The goal of this paper has been to shed some light on the proper interpretation of several response error statistics based on repeated survey measurements. Interpreting these statistics in terms of a misclassification error model demonstrated that for zero-one variables, there could be no response or measurement variance without associated measurement bias. The first summary statistic explored was the Net Difference Rate (NDR) representing the difference between survey means based on the same sample interviewed on two separate occasions. Allowing each sample member to have a distinct probability of misclassification which could differ on the two measurement occasions lead to an interpretation of NDR as the difference between average misclassification probabilities for the two measurements. Specifically, the expected value of NDR over repeated samples and replicated measurements is

$$E(\text{NDR}) = \pi[\bar{P}_G(0|1) - \bar{P}_{G'}(0|1)] + (1 - \pi)[\bar{P}_G(1|0) - \bar{P}_{G'}(1|0)] \quad (5.1)$$

where π represents the true proportion of population members with the attribute in question and $\bar{P}_G(1 - s|s)$ depicts the average probability that population members with true value $s = 1$ or 0 will respond $1 - s$ on occasion G . The more time that elapses between the measurement occasions G and G' , the more one might expect the average misclassification probabilities in 5.1 to differ.

The other two measurement error statistics examined, namely, the "Gross Difference Rate" (GDR) and the "Index of Inconsistency" (I) are generally interpreted as measures of the simple uncorrelated measurement or response variance component. For this interpretation to be correct requires that the probability of misclassification is the same on both trials for all sample members and, furthermore, that the response errors on the two occasions are uncorrelated. Section 4 of this report demonstrates that $POA = 1 - GDR$ and $\hat{R} = 1 - \hat{I}$, the complementary "Probability of Agreement" and "Index of Reliability" statistics, should be viewed as measures of correlation between the associated first and second trial response errors when the misclassification probabilities have not changed significantly. When NDR deviates significantly from zero, the interpretation of GDR and $\hat{I}$ are difficult. For cases where the misclassification process is likely to have change significantly between the two measurement occasions, a measure of association which does not rely on $NDR = 0$ should be used. The product moment correlation for bivariate binomial data would be appropriate, or possibly one of the other measures of association for two-way tables summarized by Kendal and Stuart [4] such as Cramér's C. For the 2×2 table resulting from repeated binary responses, Cramér's C is equivalent to the product moment correlation. The 1975 Statistical Package for the Social Sciences (SPSS) manual refers to the statistic above as Cramér's V.

REFERENCES

- [1] Hansen, M. R., W. N. Hurwitz, and M. A. Bershad. "Measurement Errors in Censuses and Surveys." Bull. International Statistical Institute, 38 (1961), 359-374.
- [2] Hansen, M. H., W. N. Hurwitz, and L. Pritzker. "The Estimation and Interpretation of Gross Differences and Simple Response Variance." Contributions to Statistics. Oxford, England: Pergamon Press, 1964.
- [3] Bhapkar, V. P. "A Test for Homogeneity of the Marginal Distributions in a Two-Way Classification." Biometrika: 41(1954), 412-416.
- [4] Kendal, M. C. and A. Stuart. The Advanced Theory of Statistics. Vol. 2, 2nd Ed. New York: Hafner Publishing Company, 1967, 556-566.